

# X-RAY

Steve Hammond, (National Renewable Energy Lab, steven.hammond@nrel.gov)  
Jim Laros, (Sandia National Laboratories)

## *Motivation*

It is widely recognized that emerging constraints on energy consumption will have pervasive effects on high performance computing (HPC). Power and energy consumption must now be added to the traditional goals of algorithm design, correctness, and performance. In fact, power consumption may be the single largest hardware research challenge when developing exascale systems. The Institute of Advanced Architectures and Algorithms (IAA) published a recent report in the International Journal of Distributed Systems and Technologies [a], in which it proclaimed, “The architectural challenges for reaching exascale are dominated by power, memory, interconnection networks and resilience.” Energy consumption and power management played a prominent role in the Kogge report [b] and has also been identified as a key, if not the primary, challenge for exascale systems by the Exascale Operating Systems and Runtime Technical Council. In essence, power should be viewed as a first class resource to be managed by the operating system and runtime system.

## *Approach*

To address the critical power demands of next generation platforms it is essential to understand how and where power is being used. It is also vital to determine an application’s energy profile and how this varies under different operating conditions. While individual efforts have targeted facility or platform level measurements with various degrees of success, a holistic and detailed multi-scale view is required to understand power usage in HPC environments. Ultimately, this understanding needs to be integrated with the operating system and runtime system to monitor, manage, and improve HPC energy usage and efficiency. Preliminary work by Laros [c] has already demonstrated that controlling CPU frequency

on a Cray XT class system can achieve significant energy savings with little or no impact on run-time performance.

We will establish the X-RAY (eXascale Rack AnaLYsis) laboratory for research in exascale system power and energy efficiency. The X-RAY laboratory will leverage the new ultra energy efficient HPC data center at the National Renewable Energy Laboratory (NREL). This new showcase facility is designed to be the world’s most energy efficient, featuring warm water liquid cooling, closely integrated HPC systems and building automation management, and will demonstrate power usage effectiveness (PUE) of 1.06 or better coupled with waste heat capture and reuse. The data center is instrumented to capture, evaluate and quantify system, rack and node level impacts of application and X-stack prototypes, including runtime, programming model, and domain specific libraries and languages.

As part of the X-RAY project, we will develop a software “x-ray” system with virtual probes, execution traces, and data collection that will correlate application execution patterns with power and cooling levels to drive algorithm development, code design and multilevel power management abstractions, spanning all layers of the exascale application stack.

Through the X-RAY project we will research tools, counters, and controls necessary (at the component, node, rack, system, and facility level) to expose, capture, and report power and energy. We will research ways to make node-level power management accessible to runtime system, job queues, and scheduler.

Hardware power management decisions may be relegated to the OS, which may, in turn, pass these to the runtime system, which may, in turn, pass these to the application.

Applications typically use hardware components in a predictable and controllable

time-varying profile. X-RAY will allow one to monitor and manage peak power and thermal loads in multi-level parallel architectures by explicitly setting specific compute resources in a low-power “idle” state when not being utilized. X-RAY will provide an experimental research environment where researchers can compare and contrast, and guide the development of integrated power sensitive X-stack software, including the operating system and run-time environment.

NREL’s new energy efficient HPC data will be completed this fall. The facility will enable state of the art monitoring of many metrics including temperature and power. Sandia National Laboratories has focused on component, rack and platform level power measurement and analysis with the goal of affecting HPC application performance/energy efficiency. In collaboration, we will produce a complete view of where power is being used and how power use can be managed, controlled, reduced or conserved to produce the efficiencies that we believe will be necessary for affordable Exascale computing.

Fine-grained component-level measurements are important when attempting to use runtime system software to control energy utilization at the component or sub-component level, e.g., processors/cores, NICs (Network Interface Chip/Card), and memory controllers. Typical software experiments can include parameter-variation style runs that likely need fine-grained metrics on the component being studied. Sensitivity analysis cannot be adequately done using facility, platform or rack-level measurements. For example, it is an open question as to whether or not the operating system and runtime software can ensure that power consumption never exceeds a certain upper limit.

While what is discussed here is initially specific to the capabilities being developed at NREL, the ideas and concepts are readily

generalizable and every attempt will be made to ensure portability.

- **Challenges addressed:** The X-RAY project provides a holistic approach to the power challenge, integrating how power and energy are monitored, managed, and controlled in a hierarchical system.
- **Maturity:** While this approach is in its early stages, we have very high confidence that this will be successful based on the preliminary work already done at Sandia [c].
- **Uniqueness:** The proposed approach is not unique to Exascale systems. Even petascale systems will yield direct benefits. The work could be addressed by other research programs. However, if it is not resolved, the impacts will be more pronounced at the Exascale.
- **Novelty:** This approach is unique in that it provides a holistic, integrated approach leveraging a unique new heavily instrumented data center at NREL.
- **Applicability:** As mentioned above, the proposed approach is not unique to Exascale systems. Even petascale systems will yield direct benefits.
- **Effort:** The level of effort required is probably best thought of as a collaboration with 2-3 national labs, plus university and industry partners.

#### References

- [a] K. Alvin *et al.*, “On the Path to Exascale,” *International Journal of Distributed Systems and Technologies*, April-June 2010, **1**: 1-22.
- [b] “*Exascale Computing Study: Technology Challenges in Achieving Exascale Systems*,” Peter Kogge, Editor, report from DARPA Exascale Study Group, September 2008.
- [c] Laros, J., Pedretti, K. Kelly, S., Shu, W., and Vaughan, C., “Energy Based Performance Tuning for Large Scale High Performance Computing Systems,” in preparation, 2012.